

Applicant : Wolfgang Stephan  
Serial No. : 10/601,258  
Filed : June 19, 2003  
Page : 13 of 20

Attorney's Docket No.: 13911-080001 / 2003P00147  
US

Amendments to the Drawings:

The Examiner indicated in the office action that the drawings filed 06/19/2003 (the filing date of the application) were accepted. However, the applicant notes those were informal drawings and the applicant had later filed formal drawings on September 24, 2003. For the Examiner's ease of reference, the applicant is hereby resubmitting the formal drawings. The attached replacement sheets of drawings are formal drawings replacing the informal drawings filed with the application and replace the original sheets including Figs. 1-8B.

Attachments following last page of this Amendment:

Copy of Transmittal of Formal Drawings and 7 Sheets of Formal Drawings filed on September 24, 2003.

## REMARKS

Claims 1-60 are pending. Claims 1-3, 5, 15-17, 19, 25-27, 29, 37-39, 41, 51-53 and 55 are amended. No new matter is added. Claims 8, 12, 31, 44 and 48 are rejected under 35 U.S.C. 112, second paragraph. Claims 1-30, 32-43 and 45-60 are rejected under 35 U.S.C. 102(e) as being anticipated by U.S. Patent Application Pub. No. 2004/0243560 A1 (“Broder”). The applicant respectfully traverses the rejections and requests reconsideration in view of the amendments and remarks herein.

### **I. The § 112 Rejections**

The Examiner rejected claims 8, 12, 31, 44 and 48 under 35 U.S.C. 112, second paragraph as being indefinite for failing to particularly point out and distinctly claim the subject matter which the applicant regards as the invention. In particular, the Examiner states that claims 8, 31 and 44 recite the limitation “x is in the range of 256 to 512 and y is in the range of 128 to 256”, and asserts it is unclear what type of units is involved in such range. The Examiner further asserts that the lowest range value of x “256” is the same as the highest range value of y “256”, and it is unclear whether this is a single range which covers both values of x and y or not.

Claim 8 depends from claim 1, claim 31 depends from claim 25 and claim 44 depends from claim 37. Claims 1, 25 and 37 have been amended to indicate that x and y are integers. Accordingly, the Examiner’s question as to what type of units is involved is addressed. With respect to the ranges of x and y, they are independent of one another, other than y is not equal to x. The value of x is clearly indicated in the independent claims 1, 25 and 37 as “a value x such that at least a majority of the index terms occur in x documents or fewer”. As an illustrative example, consider a collection of 500 documents and 5 index terms. If the value of x is 260, then at least 3 of the 5 index terms (*i.e.*, a “majority” of the index terms) occur in 260 documents or fewer. Claim 8 defines, for one implementation, a range for the value of x being the range of 256 to 512.

The value of y is clearly indicated in the independent claims 1, 25 and 37 as the interval between skip entries, *i.e.*, “one or more skip entries thereafter at intervals of every y<sup>th</sup> posting”. Referring again to the illustrative example above, if the value y is 140, then after the 260<sup>th</sup>

posting (*i.e.*, the  $x^{\text{th}}$  posting) there is a skip entry and thereafter there are skip entries at intervals of every  $140^{\text{th}}$  posting (*i.e.*, every  $y^{\text{th}}$  posting). The applicant respectfully submits that the range values given in claims 8, 31 and 44 are clear, and there is no indefiniteness merely because the ranges both include the integer 256.

## II. The 102 Rejections

### Claims 1-14

Claim 1 recites a method for indexing documents in a collection of document, each document including one or more index terms. Claim 1 reads as follows:

A method for indexing documents in a collection of documents, each document comprising one or more index terms, the method comprising:

determining a value  $x$  such that at least a majority of the index terms occur in  $x$  documents or fewer, where  $x$  is an integer;

determining a value  $y$ , where  $y$  is not equal to  $x$ , where  $y$  is an integer;

generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds  $x$ , further including a skip entry after the  $x^{\text{th}}$  posting and one or more skip entries thereafter at intervals of every  $y^{\text{th}}$  posting;

wherein:

a posting includes a document identifier identifying a document in the collection of documents;

a skip entry includes a document identifier that is included in a boundary posting of a block of postings immediately adjacent to the skip entry in the inverted list, where a block of postings includes postings having document identifiers ranging from a lower to an upper value and where a boundary posting is a posting having a document identifier of either the lower or the upper value.

The Examiner asserts that Broder discloses “determining a value  $x$ ” at paragraphs 0307 and 0317. The applicant respectfully submits that the Examiner has misconstrued Broder. At paragraphs 307 and 317, Broder discloses an algorithm for use with a search query where “a user wishes to retrieve the top  $n$  scoring documents for a given query. The algorithm “maintains a heap of size  $n$  to keep track of the top  $n$  results”. A document is considered a “top result” depending on a score for the document using a scoring model. For example, if there were 500 documents and a given query, each document can be given a score based on a scoring model and

the top “n” documents, *i.e.*, documents with the best score, are included in the “heap”. The value “n” is entirely different than the value “x” recited in claim 1. As discussed above in the earlier example, if there are 500 documents in a collection of documents, the majority of the index terms will occur in x documents or fewer, *e.g.*, if x equals 260, then the majority of the index terms will only occur in 260 of the 500 documents, or fewer. The value x is not a score, nor is it the number of documents in “heap” as disclosed in Broder. Accordingly, the paragraphs referred to in Broder by the Examiner fail to disclose the step of “determining a value x such that at least a majority of the index terms occur in x documents or fewer” as required by claim 1.

The Examiner asserts that Broder discloses the “generating an inverted index” step of claim 1, and in particular that Broder discloses that “if the number of postings exceeds x, further including a skip entry after the x<sup>th</sup> posting and one or more skip entries thereafter at intervals of every y<sup>th</sup> posting”, as required by the claim. The Examiner refers to Broder at paragraphs 0245 and 0277. The applicant respectfully submits the Examiner has misconstrued Broder. The applicant has amended claim 1 to clarify the definition of a skip entry, and asserts that Broder does not disclose including such skip entries in an inverted list for an index term.

Each posting in an inverted list includes a document identifier that identifies a document in the collection of documents (see claim 1, lines 11-12). A skip entry includes a document identifier that is the same as a document identifier that is included in a “boundary posting” of a block of postings immediately adjacent to the skip entry in the inverted list (claim 1, lines 13,14). A block of postings includes postings having document identifiers ranging from a lower to an upper value (claim 1, lines 14, 15). A boundary posting is a posting including a document identifier of either the lower or the upper value (claim 1, lines 15,16).

By way of illustrative example, the applicant refers to FIG. 5B in the application and paragraphs 0043 and 0044. In this example, element 510 represents a “block of postings”. A posting P<sub>257</sub> including a document identifier 303 is a boundary posting having a document identifier with the lowest value in the block 510. A posting P<sub>384</sub> including a document identifier 514 is also a boundary posting having a document identifier with the highest value in the block 510. A skip entry S<sub>1</sub> is included in the inverted list and is immediately adjacent to the block of

the postings 510. The skip entry  $S_1$  includes a document identifier of 303 (see Table 1, at para 0043). The document identifier 303 included in the skip entry  $S_1$  is the same as the document identifier included in a boundary posting (*i.e.*,  $P_{257}$ ) included in a block of postings (*i.e.*, block 510) immediately adjacent to the skip entry  $S_1$ .

The Examiner asserts that Broder discloses a skip entry at paragraph 0277, and specifically states that “there is a skip entry at the  $k^{\text{th}}$  value” [OA, footnote 2]. First, in Broder at paragraphs 0244 and 0245, a traditional inverted index for which every index term is associated with a posting list is disclosed. The list includes an entry for each document in the collection that contains the index term and each entry includes the document’s unique positive identifier, DID [see 0244]. An iterator object that is capable of sequentially iterating over a posting list is described. The iterator can skip to a given entry in the posting list. A method is described, *i.e.*, method `next(id)` that returns the first posting element for which  $\text{DID} \geq \text{id}$ . If there is no such document, the term iterator returns a special posting element with an identifier `LastID` that is larger than all existing DIDs in the index [see 0245].

Broder is thereby disclosing an iterator object that can “skip” to a given entry in an inverted list and if the entry does not exist, then a special posting element is returned. However, there is no disclosure of a “skip entry” *per se*, as the term skip entry is defined and used in the applicant’s claim 1.

In Broder’s paragraph 0277, there is a discussion of an index “ $k$ ”. The Examiner asserts: “The cursor is advanced to the position of  $k$  value. There is a skip entry at the  $k^{\text{th}}$  value.” The applicant respectfully submits that there is no disclosure in Broder’s paragraph 0277 that there is a skip entry at the  $k$ th value. A skip entry includes a document identifier that is included in a boundary posting of a block of postings immediately adjacent to the skip entry in the inverted list, as stated in claim 1. Broder does not disclose any such skip entry included in an inverted list, whether at the  $k$ th value or otherwise. Accordingly, the third limitation of claim 1 is also not disclosed by Broder.

“A claim is anticipated only if each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference” (*Verdegaal Bros. v.*

*Union Oil Co. of California*, 814 F.2d 628 (Fed. Cir. 1987)). The applicant respectfully submits the Broder reference does not anticipate claim 1. For the reasons discussed above, Broder fails to disclose at least the first and third limitations of claim 1. Accordingly, claim 1 is in condition for allowance. Claims 2-14 depend from claim 1 and are therefore allowable for at least the same reasons.

Claims 15-24

Claim 15 recites a method for indexing documents. The method includes the limitation “determining a value x, wherein at least a majority of the index terms occur in x documents or fewer and x is an integer in the range of 256 to 512”. The method further includes the limitation of “generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x<sup>th</sup> posting and one or more skip entries thereafter at intervals of every y<sup>th</sup> posting”. For at least the reasons discussed above in reference to claim 1, Broder fails to disclose the above limitations of claim 15.

Accordingly, claim 15 is not anticipated by Broder and is in condition for allowance. Claims 16-24 depend from claim 15 and are therefore allowable for at least the same reasons.

Claims 25-36

Claim 25 recites an inverted index for a collection of documents, each document including one or more index terms. The inverted index includes “one or more inverted lists including a quantity of postings that exceeds a value x, a skip entry after the x<sup>th</sup> posting, and one or more additional skip entries thereafter at intervals of every y<sup>th</sup> posting, where the value x is such that at least a majority of the index terms occur in x documents or fewer”. For at least the reasons discussed above in reference to claim 1, Broder fails to disclose an inverted list including skip entries. Accordingly, claim 25 is not anticipated by Broder and is in condition for allowance. Claims 26-36 depend from claim 25 and are therefore allowable for at least the same reasons.

Claims 37-50

Claim 37 recites an article comprising a machine-readable medium storing instructions operable to cause one or more machines to perform operations. The operations include “determining a value x such that at least a majority of the index terms occur in x documents or fewer” and “generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x<sup>th</sup> posting and one or more skip entries thereafter at intervals of every y<sup>th</sup> posting”. For at least the reasons discussed above in reference to claim 1, Broder fails to disclose the above limitations of claim 37. Accordingly, claim 37 is not anticipated by Broder and is in condition for allowance. Claims 38-50 depend from claim 37 and are therefore allowable for at least the same reasons.

Claims 51-60

Claim 51 recites an article comprising a machine-readable medium storing instructions operable to cause one or more machines to perform operations. The operations include “determining a value x, wherein at least a majority of the index terms occur in x documents or fewer and x is an integer in the range of 256 to 512” and “generating an inverted index for the collection of documents, the inverted index including an inverted list for each of the index terms, each inverted list including at least one posting and, if the number of postings exceeds x, further including a skip entry after the x<sup>th</sup> posting and one or more skip entries thereafter at intervals of every y<sup>th</sup> posting”. For at least the reasons discussed above in reference to claim 1, Broder fails to disclose the above limitations of claim 51. Accordingly, claim 51 is not anticipated by Broder and is in condition for allowance. Claims 52-60 depend from claim 51 and are therefore allowable for at least the same reasons.

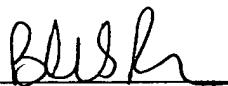
No fees are believed due, however, please apply any charges or credits to deposit account 06-1050.

Applicant : Wolfgang Stephan  
Serial No. : 10/601,258  
Filed : June 19, 2003  
Page : 20 of 20

Attorney's Docket No.: 13911-080001 / 2003P00147  
US

Respectfully submitted,

Date: Mar 27/04



Brenda M. Leeds Binder  
Reg. No. 57,520

Fish & Richardson P.C.  
500 Arguello Street, Suite 500  
Redwood City, California 94063  
Telephone: (650) 839-5070  
Facsimile: (650) 839-5071

50325584